Complexity of Stochastic Optimization

Benjamin Armbruster

Department of Management Science and Engineering
Stanford University

March 9, 2006 / Stanford INFORMS Seminar

Outline

Stochastic Optimization

$$
\begin{aligned}
(z, x^*) = \min_x \; & f(x) := \mathbb{E}_\omega \, F(x,\omega) := \int_\Omega F(x,\omega) \, \mathrm{d}P(\omega) \\
s.t. \; & x \in K
\end{aligned} \tag{1}
$$

- $K \subseteq \mathbb{R}^d$
- K and F are "nice"

- K convex
- F convex $\implies$ f convex

- we can sample from P

Ellipsoid Method

$$\min_x f(x) \qquad (2)$$
$$s.t.\ x \in K := \{x \ : \ g(x) \leq 0\}$$

Assumptions

- f convex
- g convex $\implies K \subseteq \mathbb{R}^d$ convex
- oracles for f, f', g, g'
- $B(r) \subseteq K \subseteq B(R)$

Result: $f(x_N) - z^* \leq \epsilon$ and $g(x_N) \leq \epsilon\vec{1}$ if

$$N = O(1)d^2 \log \frac{R \max |f|}{r\epsilon}$$

Gradient Descent

$$\min_{x \in K} f(x) \tag{3}$$

Assumptions:

- f convex
- L Lipschitz constant of f
- oracle for f'
- ignoring constraints for now
- $K \subseteq B(R)$

Algorithm:

$$x_{i+1} = x_i - \alpha_i \frac{f'(x_i)}{\|f'(x_i)\|}$$

Step size: $\alpha_i = R/\sqrt{i}$
Performance:

$$f(x_N) - z^* \leq \epsilon \quad \text{if} \quad N = O(1)\frac{L^2 R^2}{\epsilon^2}$$

Stochastic Approximation

Assumptions:

- $L := \sqrt{\sup_{x \in K} \mathbb{E} \| F'(x, \omega) \|^2}$
- oracle for F' (and assume $f' = \mathbb{E} F'$)
- ignoring constraints for now
- $K \subseteq B(R)$

Algorithm:

- i.i.d. samples $\omega^1, \ldots$
- $y_{i+1} := y_i - \alpha_i F'(y_i, \omega_i)$
- step size: $\alpha_i := \frac{R}{L\sqrt{i}}$
- $x_N := \left[\sum_{N/2}^{N} \alpha_i \right]^{-1} \sum_{N/2}^{N} \alpha_i y_i$

Performance:

$$\mathbb{E} f(x_N) - z \leq \epsilon \quad \text{if} \quad N = O(1) \frac{L^2 R^2}{\epsilon^2}$$

Improved Convergence

- Assume $c_1 \|x - x^*\|^2 \leq f(x) - z \leq c_2 \|x - x^*\|^2$
- $y_{i+1} := y_i - \alpha_i F'(y_i, \omega_i)$
- step size: $\alpha_i := \frac{3}{i c_1}$
- $x_N := \left[\sum_{N/2}^{N} \alpha_i\right]^{-1} \sum_{N/2}^{N} \alpha_i y_i$

Performance:

$$\mathbb{E} f(x_N) - z \leq \epsilon \quad \text{if} \quad N = O(1) c_2 \frac{R^2 + L^2/c_1^2}{\epsilon}$$

Sample Average Method

original problem $(z, x^*) = \min_{x \in K} f(x) := \mathbb{E}_\omega F(x, \omega)$

sample average problem take i.i.d. samples $\omega^1, \dots, \omega^N$

$$(z_N, x_N) = \min_x \; f_N(x; \omega^1, \dots, \omega^N) := \frac{1}{N} \sum_{i=1}^{N} F(x, \omega^i)$$
$$s.t. \; x \in K \tag{4}$$

generally $z_N \to z$, $x_N \to x^*$, and $f(x_N) \to z$

Known Results for Sample Average Method

- $\sqrt{N}(x_N - x^*) \Rightarrow X$
- $\Pr[f(x_N) - z \geq \epsilon] \leq \delta$ if

$$N \geq \frac{O(1)\widehat{R}^2}{\epsilon^2}\left[d \log \frac{2RL}{\epsilon} + O(1) + \log \frac{1}{\delta}\right] \tag{5}$$

  Assuming
  - $F(\cdot, \omega)$ Lipschitz with modulus L
  - $K \subset \mathbb{R}^d$ of diameter R
  - $\widehat{R} := \sup_{x,\omega} |F(x, \omega)|$

New Results

The program:

1. Assume $c_1 \left\|x - x^*\right\|^2 \leq f(x) - z \leq c_2 \left\|x - x^*\right\|^2$, then

$$\Pr[\|x^* - x_N\| \geq \sqrt{\epsilon/c_1}] \leq \Pr[f(x_N) - z \geq \epsilon] \leq \Pr[\|x^* - x_N\| \geq \sqrt{\epsilon/c_2}]$$

2. $\Pr[\|x^* - x_N\| \geq \Delta] \leq \delta$ if

$$N \geq \frac{2C_0^2}{\Delta^2 c_1^2} \left[\log \frac{1}{\delta} + d \log \log \frac{1}{\delta} + O(1)\right]$$

3. So, $\Pr[f(x_N) - z \geq \epsilon] \leq \delta$ if

$$N \geq \frac{2C_0^2 c_2}{\epsilon c_1^2} \left[\log \frac{1}{\delta} + d \log \log \frac{1}{\delta} + O(1)\right]$$

$C_0 := \sup_{x,i,\omega} |F_i'(x,\omega)|$

Comparison

General complexity:

stochastic approximation: $O(1)\frac{L^2R^2}{\epsilon^2}c(\delta)$

sample average method: $O(1)\frac{\widehat{R}^2}{\epsilon^2}c(\delta)$

Assuming $c_1\left\|x-x^*\right\|^2 \leq f(x)-z \leq c_2\left\|x-x^*\right\|^2$

stochastic approximation: $O(1)c_2\frac{R^2+L^2/c_1^2}{\epsilon}c(\delta)$

sample average method: $c_2\frac{C_0^2}{\epsilon c_1^2}c(\delta)$

Two-period problems

$$\min_{x \in K} f(x) := \mathbb{E}_\omega F(x, \omega)$$

where

$$F(x, \omega) := \min_{y(\omega)} \; r(y; x, \omega)$$
$$s.t. \; g(y; x, \omega) \leq 0$$

Multi-period problems probably hard

in a T period stochastic problem

$$\min_{x \in K} f(x) := \mathbb{E}_\omega F(x, \omega)$$

F is the solution to a $T-1$ period problem
sample average method needs $O(\epsilon^{-2(T-1)})$ samples

Discussion

- how do you deal with multiple stages?
- how do you choose the probability distributions?
- ϵ^{-2} or ϵ^{-1} complexity probably good enough for real-world problems

Summary

- generally ϵ^{-2} complexity
- a $f'(x^*) = 0$ implies ϵ^{-1} complexity
- stochastic approximation and sample-average methods quite similar

References

BA2006 *Sample Complexity of Smooth Stochastic Optimization* `www.stanford.edu/~barmbrus`

Shapiro,Nemirovski 2005 *On complexity of stochastic programming problems* eprint: `www.optimization-online.org`

Dyer,Stougie 2006 *Computational complexity of stochastic programming problems* Math. Prog. 106:3